Model grid metadata

V. Balaji
Princeton University and NOAA/GFDL

GO-ESSP Meeting
Rutherford Appleton Laboratories, Chilton, UK
6 June 2005

Data consumers

Scientists perform sequences of computations (e.g ***"poleward heat transport", "meridional overturning streamfunction"***) on datasets. Typically this is scripted in some data analysis language, and ideally it should be possible to apply the script to diverse datasets.

The IPCC data archive at PCMDI has been a success for consumers without precedent, and will be cited in many groundbreaking works of climate research for many years to come.

Data producers

Observational and model output data in the climate-ocean-weather (COW) community is initially generated in native format, and any subsequent relative analyses requires considerable effort to systematise. Issues include moving and transient data sources, lossy data formats, curvilinear and other “exotic” coordinates.

Data managers

Data managers are the community within this ecosystem that facilitates the transformation of source dependent data to a neutral and readily consumable form. They develop the standards for describing data in a manner that permits these transformations, and develop tools to perform them.

The data ecosystem

We identify three communities, data ***consumers***, ***producers*** and ***managers***.

- Data is created in a manner most suited to producers (models, observations).
- It is delivered to consumers in a manner where data from different sources can be merged and coherently analysed.
- The manager niche in the ecosystem should take responsibility for mediating between these two communities. This is where CF, GO-ESSP play a role.

The key issue is to make it possible not only to display, but to construct a scientific study using, data from different sources, based on the datasets alone.

Standards play a role...

Model metadata: describing data source comprehensively, relatively easy for observations, harder for models but can asymptote toward completeness starting from current PCMDI standard

Physical fields: standard vocabulary for describing the relevant physical quantities (viz. CF `standard_name`). Variables can contain ***gridded*** or ***point*** (station, drifter) data.

Geospatial information: location information. This set of standards unites a much larger community (mobiles, GIS), in which our community has begun to play a role.

Grid metadata: interrelations between grids, between points and grids.

Grid metadata: what's included

The grid specification includes ***distances, areas, angles and volumes*** for decomposing thin spherical shells or cartesian space.

It could also contain specifications for ***exchange grids*** and ***masks***.

We apply thin-fluid scaling arguments to separate out the vertical. The vertical coordinate can be space- or mass-based.

LRGs and UPGs

LRG logically rectangular grid.

STG structured triangular grid.

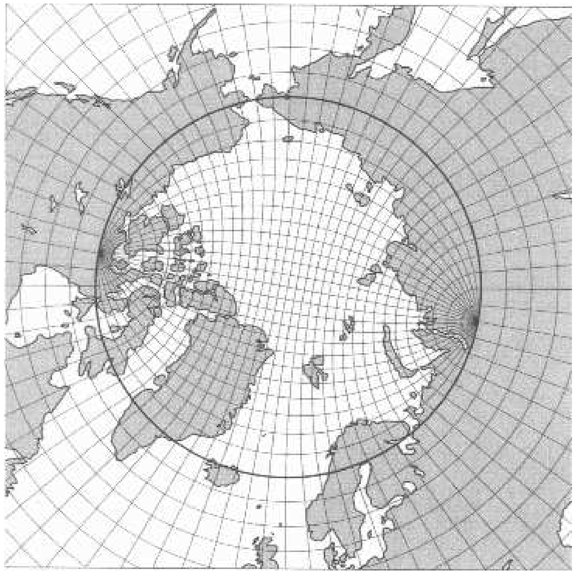
UTG unstructured triangular grid.

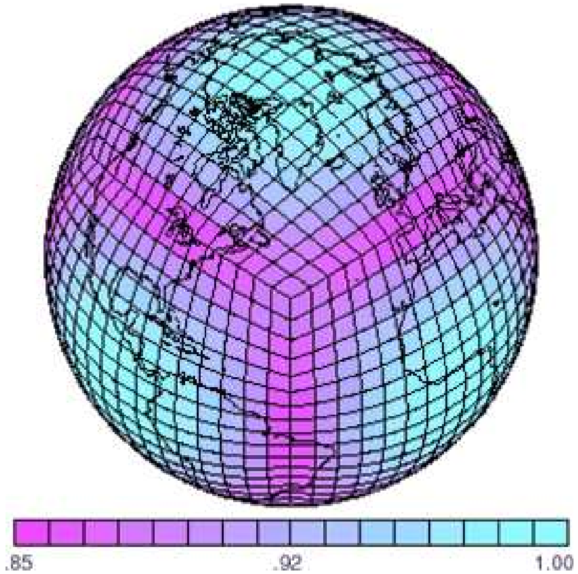
UPG unstructured polygonal grid.

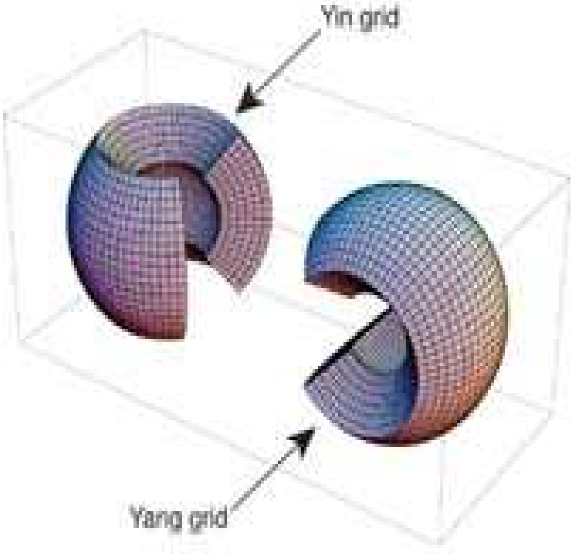
These allow us to describe all conceivable grids for the near- to mid-future.

An actual grid may be composed of a ***mosaic*** of LRGs or UPGs. In principle, you could even mix them (i.e define a grid with some LRG and some UPG tiles).

Non-”standard” LRGs



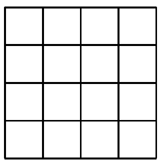




On the left is the ***tripolar grid*** (Murray 1996, Griffies et al 2004) used by MOM4 for the current IPCC model CM2. In the middle is the ***cubed sphere*** (Rancic and Purser 1990) planned for the Finite-Volume atmosphere dynamical core for the next-generation GFDL models AM3 and CM3. On the right is another promising grid, the ***yin-yang grid*** (Kageyama et al 2004).

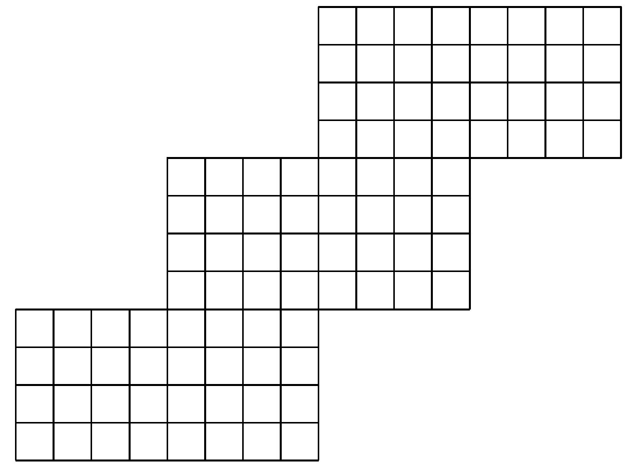
A key difference between these is that the tripolar grid is a single LRG.

What is a mosaic?

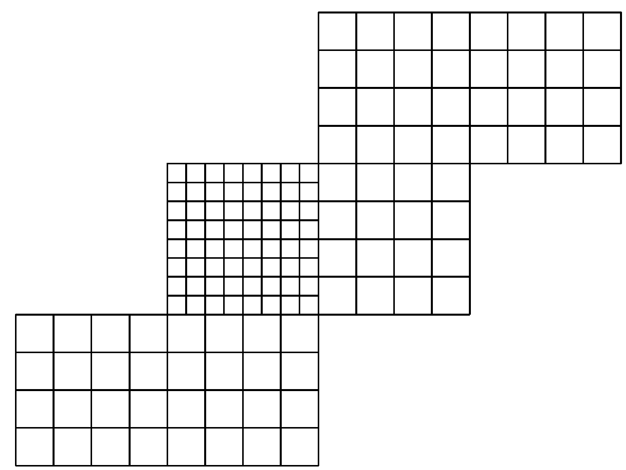


On the left is a basic 4×4 ***tile***; on the right are examples of grids composed of a mosaic of such tiles. The first is a ***continuous grid***, below is a ***refined grid***.

Most current software only supports what we call ***tiles*** here. The **`mosaic`** extension will allow the development of more complex grids for next-generation models.
First in our (GFDL's) sights is the ***cubic sphere***, primarily targeted at a next-generation finite-volume atmospheric dynamical core, but potentially others as well.
Further developments will include support for irregular tiling (e.g of the ocean surface following coastlines), and for refined, nested and adaptive grids.
Also, regular grids where an irregular decomposition is needed (e.g for a polar filter) can use mosaics to define different decompositions in different regions.

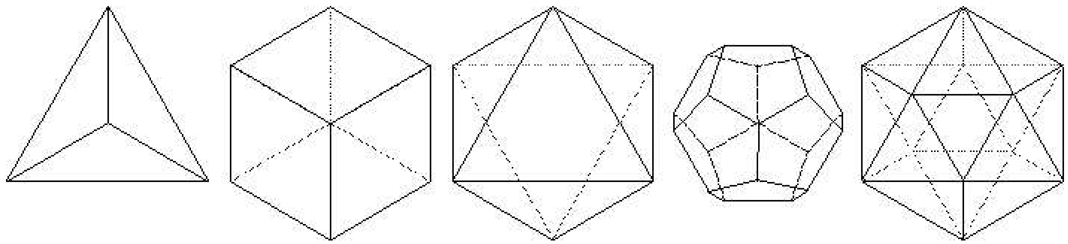


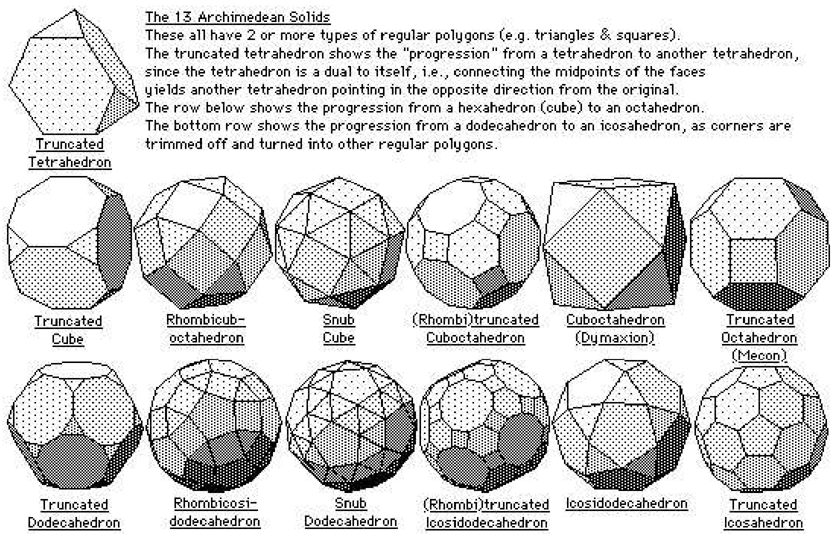
Regular grid mosaic.



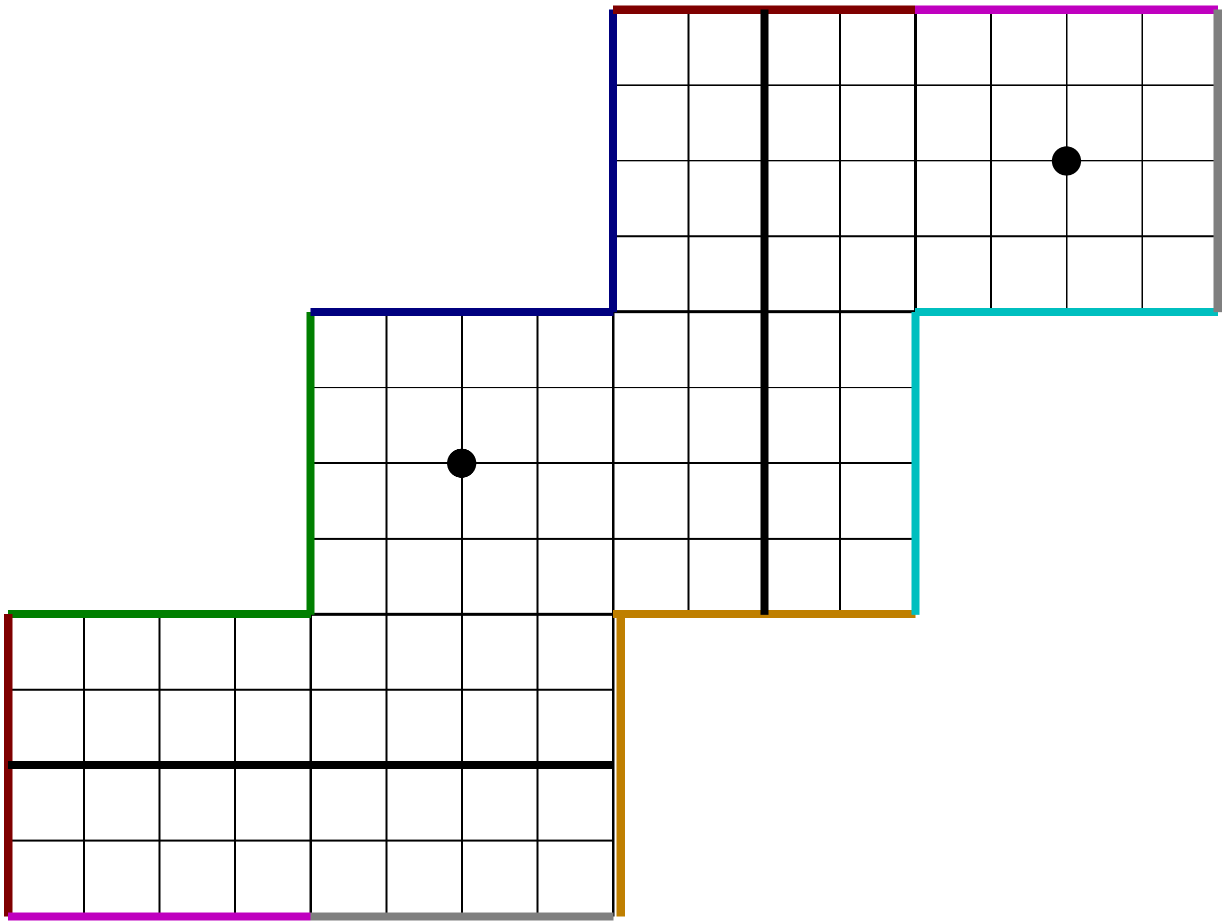
Refined grid mosaic.

UPG mosaics are likely to become more common



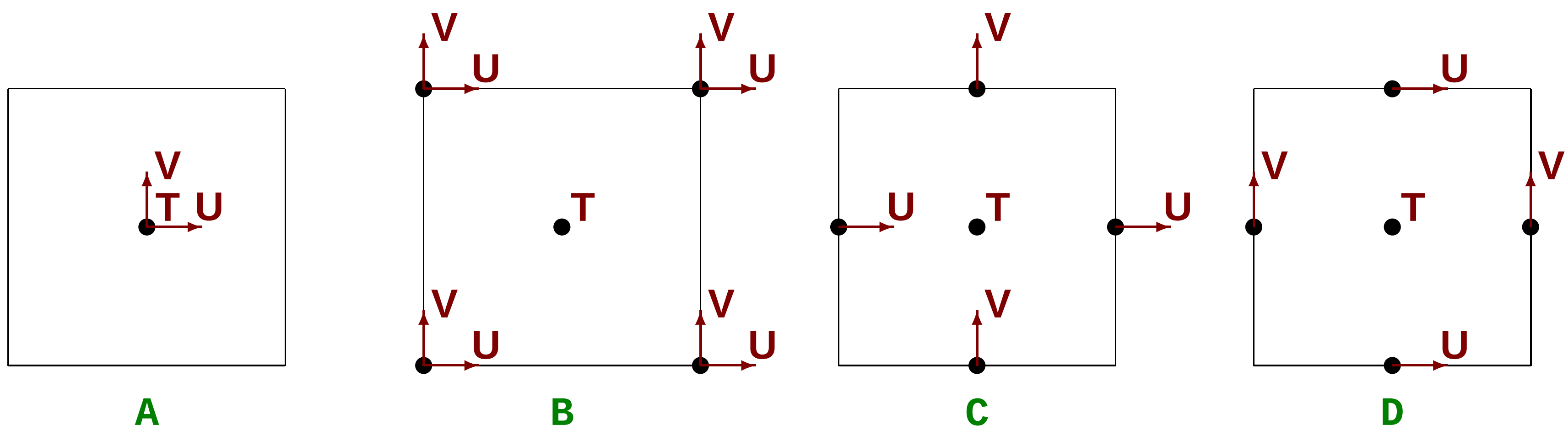


Cubed sphere



Mosaic topology for the cubed sphere. Note that boundaries may change orientation: the point just to the “west” of **`(5,6)`** is in fact **`(3,4)`**; and furthermore vector quantities transiting the boundary at that point will undergo rotation.

Supergrids

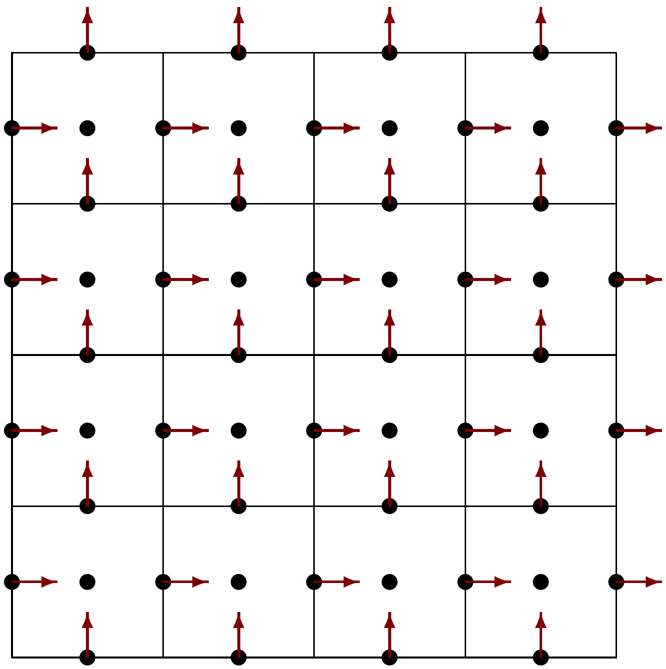


Algorithms place quantities at different locations within a grid cell (“staggering”). This has led to considerable confusion in terminology and design: are the velocity and mass grids to be constructed independently, or as aspects (“subgrids”) of a single grid? How do we encode the relationships between the subgrids, which are necessarily fixed and algorithmically essential?

In this specification, we dispense with subgrids, and instead invert the specification: we define a **supergrid**. The ***supergrid*** is an object potentially of higher refinement than the grid that an algorithm will use; but every such grid needed by an application is a subset of the supergrid.

LRG supergrids are themselves LRGs while a UPG supergrid can always be described by a unstructured triangular grid (UTG).

Specifying a single tile: C-grid LRG



The supergrid is defined as 9×9:

```
gridspec_version = 0.1
nx = 9
ny = 9
geographic_latitude(nx,ny)
geographic_longitude(nx,ny)
grid_east_angle(nx,ny)
grid_north_angle(nx,ny)
dx(nx-1,ny)
dy(nx,ny-1)
area(nx-1,ny-1)
intend_x_refinement = 2
intend_y_refinement = 2
```

(1)

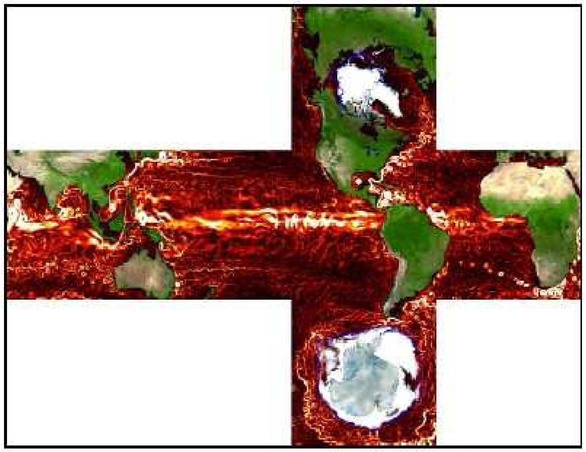
Optional keywords **`uniform`**, **`orthogonal`** are used to compress the spec.

```
gridspec = "gridspec.nc"
   checksum = "..."
nx_u = 1:nx:intend_x_refinement
ny_u = 2:ny:intend_y_refinement
grid_eastward_velocity(nx_u,ny_u)
```

(2)

From tiles to mosaics

If each tile is written out separately, current software is already capable of displaying results:



Any computation that crosses a tile boundary involves the specification of ***contact regions*** between tiles. Contact regions cannot necessarily be deduced from geospatial information.

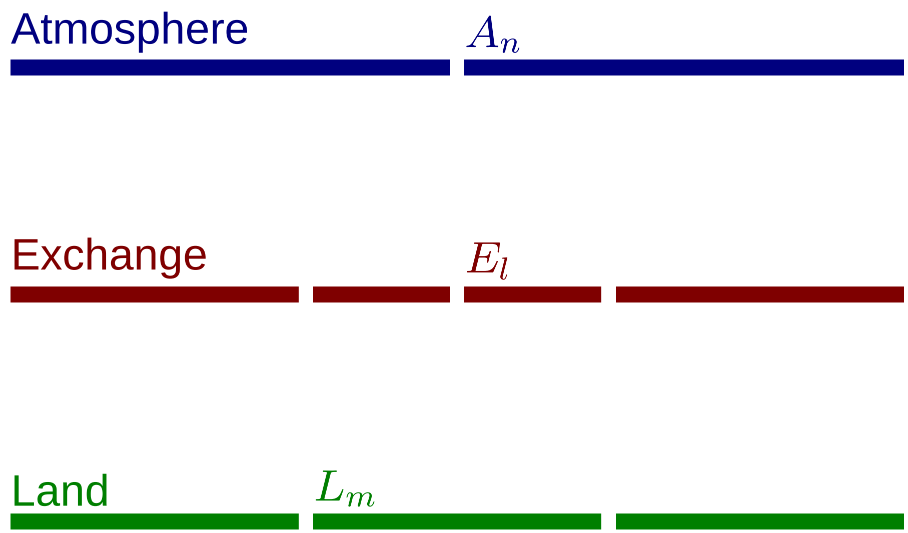
```
mosaic_version = 0.2
mosaic "atmos"
   grid_type "cubed_sphere"
   grid_mapping
   tile "face1"
   ...
mosaic "ocean"
   grid_type "tripolar_grid"
   grid_mapping
   tile "tile"
contact_region "atmos:face1" "ocean:tile"
   ncells
   parent(ncells,2)
   frac_area(ncells,2)
   mask
```

(3)

Regridding is a key operation that requires contact region information. ***Conservative*** regridding between multiple ***components*** or ***nests*** requires ***exchange*** and ***mask*** information.

Definition of an exchange grid

- A ***grid*** is defined as a set of ***cells*** created by ***edges*** joining pairs of ***vertices*** defined in a discretization.
- An ***exchange grid*** is the set of cells defined by the union of all the vertices of the two parent grids, and a ***fractional area*** with respect to the parent grid cell.



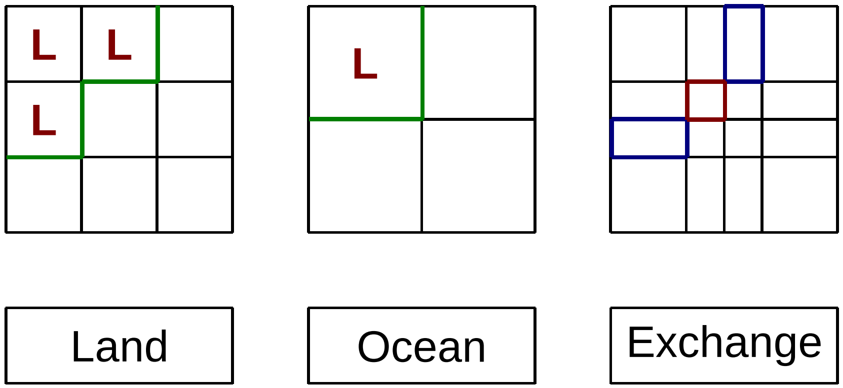
- Exchange: interpolate from source grid using one set of fractional areas; then average onto the target grid using the other set of fractional areas.

- Consistent moment-conserving interpolation and averaging functions of the fractional area may be employed.

Masks

Complementary components: in Earth system models, a typical example is that of an ocean and land surface that together tile the area under the atmosphere.

Land-sea mask as discretized on the two grids, with the cells marked **L** belonging to the land. Certain exchange grid cells have ambiguous status: the two blue cells are claimed by both land and ocean, while the orphan red cell is claimed by neither.



Therefore the mask defining the boundary between complementary grids can only be accurately defined on the exchange grid.

In the FMS exchange grid, by convention (and because it is easier) we generally modify the land grid as needed. We add cells to the land grid until there are no orphan "red" cells left on the exchange grid, then get rid of the "blue" cells by ***clipping*** the fractional areas on the land side.

Extensions to current grid specification in CF

- The specification of a tile is consistent with the current CF gridspec, but extends it by defining the supergrid and staggering.

- The definition of a mosaic is new. The mosaic specification can help widen the parallel I/O and filesize bottlenecks.

- The grid specification is maintained separately from the dataset, which links to it. Integrity of linkages between files is maintained by adding a **`checksum`** attribute to each linked file.

- If the gridspec file is standardized, it can be used for model input as well as output. For coupled or nested models, this file may also contain the necessary data to relate component grid mosaics.

What's needed next

- prototype and test across more than one institution;
- CF to agree to an extended standard for gridded datasets;
- PRISM/ESMF to agree to produce compliant data;
- Tools to become capable of applying standard and bespoke regridding techniques.
- If agreed tomorrow, GFDL will propose a draft standard, a compliant `gridspec.nc` file, and sample datafiles consistent with the gridspec.